



Water level observations from Unmanned Aerial Vehicles for improving estimates of surface water-groundwater interaction

Bandini, Filippo; Butts, Michael; Vammen Jacobsen, Torsten; Bauer-Gottwein, Peter

Published in:
Hydrological Processes

Link to article, DOI:
[10.1002/hyp.11366](https://doi.org/10.1002/hyp.11366)

Publication date:
2017

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Bandini, F., Butts, M., Vammen Jacobsen, T., & Bauer-Gottwein, P. (2017). Water level observations from Unmanned Aerial Vehicles for improving estimates of surface water-groundwater interaction. *Hydrological Processes*, 31(24), 4371-4383. <https://doi.org/10.1002/hyp.11366>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Water level observations from Unmanned Aerial Vehicles for improving estimates of surface water-groundwater interaction

Filippo Bandini^a, Michael Butts^b, Torsten Vammen Jacobsen^b, and Peter Bauer-Gottwein^a

^aDepartment of Environmental Engineering, Technical University of Denmark, 2800 Lyngby, Denmark

^bWater Resources Department, DHI, 2970 Hoersholm, Denmark.

Corresponding author: Filippo Bandini (fban@env.dtu.dk). Tel.: 0045-53336482

Running title: UAV-borne water level monitoring for hydrological model calibration

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hyp.11366

Abstract

Integrated hydrological models are usually calibrated against observations of river discharge and piezometric head in groundwater aquifers. Calibration of such models against spatially distributed observations of river water level can potentially improve their reliability and predictive skill. However, traditional river gauging stations are normally spaced too far apart to capture spatial patterns in the water surface, while spaceborne observations have limited spatial and temporal resolution. UAVs (Unmanned Aerial Vehicles) can retrieve river water level measurements, providing: i) high spatial resolution; ii) spatially continuous profiles along or across the water body; iii) flexible timing of sampling. A semi-synthetic study was conducted to analyse the value of the new UAV-borne datatype for improving hydrological models, in particular estimates of GW (Groundwater)- SW (Surface Water) interaction. Mølleåen River (Denmark) and its catchment were simulated using an integrated hydrological model (MIKE 11-MIKE SHE). Calibration against distributed surface water levels using the DiffereNtial Evolution Adaptive Metropolis (DREAM) algorithm demonstrated a significant improvement in estimating spatial patterns and time series of GW-SW interaction. After water level calibration, the sharpness of the estimates of GW-SW time series improves of ca. 50% and RMSE (Root Mean Square Error) decreases by ca. 75% compared to a model calibrated against discharge only.

Keywords: UAV, water level, groundwater surface water interaction, MIKE 11/ SHE, radar altimetry, DREAM algorithm

Key Points:

- Spatially distributed UAV-borne water level observations are an innovative dataset in hydrology
- A river model was calibrated against UAV-borne water level observations using the DREAM algorithm
- Calibration against distributed water levels improved estimates of surface water groundwater interaction

1. Introduction

Calibration and validation of integrated hydrological models have traditionally focused on water balances, groundwater levels, and river discharge. However, hydraulic modelling of water levels within river channels and floodplains is essential to capture interactions between the river and its surroundings. Indeed, many local features and processes (e.g., channel bathymetry, vegetation, interaction of river with floodplains, river control structures) strongly influence water levels while only to a lesser extent river discharge (Yamazaki *et al.*, 2012). For this reason, spatial coverage, accuracy and resolution of water level observations are essential for improving surface water management and flood prediction (Yan *et al.*, 2015). Historically, water level and discharge measurements have been the backbone for hydrological modelling, and, in the last 20 years, remote sensing has contributed hydrological observations with improved spatial and temporal coverage (Alsdorf *et al.*, 2007). However, obtaining comprehensive observational datasets of water level in medium and small sized rivers with optimal accuracy and spatial resolution is still an unresolved challenge.

1.1. In-situ measurements of water level

River gauging networks are generally sparse. Discharge is often estimated from gauged water levels through rating curves; however, observed water levels are not recorded in a worldwide consistent dataset (Yamazaki *et al.*, 2012). Furthermore, the number of operational and accessible river gauging stations has been decreasing in the last decade (Lawford *et al.*, 2013).

1.2. Spaceborne measurements of water level

Satellite altimetry data are generally processed and optimized for sea water levels and thus not useful for monitoring small continental water bodies (Legresy *et al.*, 2005; Schumann and Domeneghetti, 2016). Thus radar altimetry has limited spatial resolution and may not always reliably map rivers that are less than 1 km wide (Domeneghetti *et al.*, 2015). Indeed, popular satellite altimetry sensors, such as those on board Envisat, Topex, ERS2, Jason I and II, have ground footprints of several hundreds of meters, thus can accurately monitor water bodies only when their width is larger than the footprint (O'Loughlin *et al.*, 2016). However, water levels of medium-large size rivers (width between 100 and 1000 m) can be identified by incorporating a priori information such as the exact location, width, and shape of the river in the waveform analysis (Maillard *et al.*, 2015).

The new generation radar altimetry such as Synthetic Aperture Interferometric Radar Altimeter (SIRAL) on board CryoSat-2 has an along-track resolution of ca. 250-300 m (Wingham *et al.*, 2006). When operated in SARin mode, the use of a second antenna (interferometry) allows correction for the cross-track slope (Villadsen *et al.*, 2015). Moreover, its waveform shape, which is different from conventional altimeters, enhances separation between water and surrounding topography through novel retracking algorithms (Kleinherenbrink *et al.*, 2014). Nonetheless, observation of narrow rivers (less than 100 m wide) with an accuracy of better than 0.5 m, is still a major challenge. The Geoscience Laser

Altimeter System (GLAS), which has a footprint of 70 m and an along track distance between consecutive footprints of 170 m, has shown the possibility to retrieve water levels at decimetre accuracy (Baghdadi *et al.*, 2011; Hall *et al.*, 2012; Phan *et al.*, 2012). However, the removal of bank and vegetation contamination is still challenging for spaceborne LIDAR.

1.3. Airborne measurements of water level

Airborne LIDAR has been successful for water surface elevation measurements in narrow rivers (Schumann *et al.*, 2008). However, Digital Elevation Models (DEMs) retrieved by airborne LIDAR have limited spatial coverage and do not capture temporal dynamics. This is mainly because airborne surveys are expensive. In addition, the accuracy of the airborne LIDAR system depends on the surveying environment, e.g. size of the water surface, vegetation cover, topography, geometry (Mason *et al.*, 2007). Airborne LIDAR surveys over water showed offsets from -0.22 to +0.04 m, with an overall mean offset of ca. -0.06 m (Hopkinson *et al.*, 2011).

1.4. UAV-borne measurements of water level

Only UAVs ensure the accuracy and spatial resolution to monitor small terrestrial water bodies, including narrow rivers (less than 100 m wide). To date, few studies have analysed the potential of UAVs in retrieving water level observations. Photogrammetry is a well-known technique (Giordan *et al.*, 2016), but has low accuracy in identifying water surface (e.g. decimetre-metre level) because it is strongly influenced by water turbidity, sun and shadow conditions, vegetation, GNSS (Global Navigation Satellite System) and IMU (Inertial Measurement Unit) inaccuracies, distortion of the camera lens, etc.. To account for some of these problems the orthophoto generation requires ground control points (GCPs). A slightly different approach is proposed by Niedzielski *et al.* (2016), who intentionally omit the use of GCPs. In this case, a previous airborne LIDAR survey was used to provide a spatial fix and correct for errors during orthomosaicking of the UAV images. The authors documented that

they can observe the extent of the water surface and can classify river stages in low, normal and high-flow situations.

Bandini *et al.* (2017) adopted a different approach, documenting the possibility of measuring accurate orthometric water levels from UAVs by using a system comprising a differential GNSS system and a sensor to range water surfaces (radar).

1.5. Rationale

We conduct a semi-synthetic study to analyse the potential of the new datatype, UAV-borne river water level observations, for improving hydrological models. With our framework, we demonstrate that spatially distributed UAV-borne observations can improve knowledge of the probabilistic distribution of the river model parameters, and enhance sharpness and reliability of GW (Groundwater) - SW (Surface Water) interaction estimates. River–aquifer exchange flow is governed by the head difference between the two systems, by the river geometry and by the hydraulic conductivity (Munz *et al.*, 2011). Therefore, accurate observations of the river water level are essential to determine the direction and magnitude of the exchange flow. For this case study, estimating GW-SW exchange flow is important to evaluate impacts of e.g. groundwater abstraction on streamflow. Indeed, as demand for groundwater increases and groundwater levels decrease around pumping wells, the resulting hydraulic gradients, can draw water from the surface water body into the aquifer (“stream depletion”). Furthermore, pump-induced drawdowns in the aquifers can make the direction of the exchange flow variable at short spatial scales.

2. Materials and methods

First, we show the potential of UAV-borne water level monitoring at specific locations of Mølleåen River, Denmark. Second, a synthetic study was performed to analyse the value of this new data for hydrological model calibration and prediction.

2.1. UAV-borne water level monitoring technology

Bandini et al. (2017) describe a UAV-borne system for retrieving orthometric water levels. The system comprises a GNSS system and a ranging sensor. The ranging sensor measures the range to the water surface, whilst the flight altitude above mean sea level is retrieved by the GNSS system. The difference between the two yields the elevation of the water surface. The ranging sensor consists of a 77 GHz (W band) radar sensor with a mechanical scanning antenna and a reading rate of 15 Hz. It is able to retrieve multiple targets in its field of view with a resolution of 10 cm. Accurate horizontal and vertical positions are retrieved using a carrier phase differential GNSS system with a position acquisition rate of 5 Hz. The radar is able to measure the range and angle of each of the multiple targets in its field of view. Thus, it is expected to be able to retrieve highly spatially distributed water level observations also in narrow rivers. The measuring accuracy of the integrated system, which includes the GNSS receiver and the radar sensor, is within 5-7 cm.

2.1.1. Distributed water level observations

For this case study, water level observations were retrieved in the downstream branch of Mølleåen River, as shown in Figure 1. Mølleåen is a Danish stream that has a total length of around 30 km and a width that varies from a few meters to some tens of meters.

As a proof of concept, this paper reports the actual water level observations that were retrieved by the UAV platform over some stretches of the river in Figure 1. However, in other

river stretches, a flight survey was not possible, due to both legislation and safety issues. The main issues are the difficulty to maintain constant visual contact with the flying platform (as required by current Danish legislation) and a limited number of areas where the drone is permitted to land. For this reason, to cover all the areas not surveyed by the UAV, ground-based measurements were retrieved using an RTK (Real Time Kinematic) GNSS rover station.

Figure 2 shows the water level profile of the downstream branch of the river.

Figure 2 displays the river water profile, obtained by integrating ground-based and UAV-borne water level observations. Water slope is controlled mainly by the river structures (weirs). The uncertainty of ground-based observations is due to the accuracy of the RTK rover station, which achieves a vertical accuracy of 5-6 cm under clear-sky conditions. The standard deviation of the UAV-borne observations depends on the positioning and ranging performance of the UAV-borne measuring system (Bandini *et al.*, 2017). UAV-borne water level observations retrieved in one of the stretches are presented in Figure 3, together with the flight path.

Figure 3 shows that the water level observations were retrieved by flying the drone above the river, approximately above the centre line, with a flight time of ca. 400 s. The drone flew at an altitude of ca. 30 m above ground level. The radar is able to retrieve up to 32 targets in near range field, thus the target representative of the water surface needs to be identified. The recorded target angle allows a pre-screening of the targets. Indeed, since the drone IMU records the drone pitch and roll angles, the expected angle of the water surface with respect to the drone (e.g. water at nadir angle) is known for each observation. This expected angle of the water surface is then used to predetermine the radar targets that should be representative

of the water surface (i.e. water target within $\pm 2^\circ$ of the expected angle). If multiple radar targets are in this $\pm 2^\circ$ range, the exact target is selected by taking into account spatial continuity of the water surface observations. This procedure removes peaks due to interference from the surroundings, such as vegetation canopy. Climbs and dives in the water level observations seen in Figure 3 are due to GNSS-derived position inaccuracies and to multipath distortion of the radar signal; however, the slope of the river can be clearly determined after a filter is applied to the observations. The filter consists of a centred moving average. For computation of the moving average, the river was first subdivided into 1-m intervals and all measurements falling on the same interval were averaged. Subsequently, a 50-points moving average filter was applied to the 1-m resolution water level dataset.

2.2. Analysis of data value

In the synthetic study, we investigate whether calibration of an integrated hydrological model against synthetic spatially distributed water levels improves reliability and sharpness of GW-SW exchange flow estimates.

2.2.1. Model setup

Mølleåen River was simulated together with its hydrological catchment with the hydrological model MIKE 11-MIKE SHE (Sole and Zuccaro, 2003; Graham and Butts, 2005; Hughes and Liu, 2008). Mølleåen river was chosen for this study because the exchange flow direction varies both spatially and temporally, especially due to pumping-induced hydraulic gradients. The catchment area is shown in Figure 4: it comprised Mølleåen, the main river included in the model, some large lakes (Lyngby Sø, Bagsvaerd Sø, Furesø, Farum Sø, Vejlesø and Søllerød Sø) and other minor creeks. The hydrodynamic river model is informed with in-situ observed river cross-sections at spatial intervals of hundreds of meters. The integrated

hydrological model was calibrated against the piezometric head and discharge observations shown in Figure 4.

We evaluate the potential of spatially distributed water level observations for the branch flowing from Lyngby Sø to the sea, by comparing two different calibrations of the river model: i) calibration only against discharge observations, which we will refer to as discharge calibration (DC), and ii) calibration against discharge and spatially distributed water level observations, which we will refer to as discharge and water level calibration (DWLC). The simulation period consisted of a warm-up period of around 5 years to reduce sensitivity to initial conditions, a calibration period of the river model of ca. 3 months and an evaluation period of ca. 3 months.

2.2.2. *Synthetic observations*

Spatially distributed measurements of GW-SW interaction were not available. However, observations of GW-SW flow were necessary for evaluating the improvement on GW-SW estimates after model calibration against water level observations. Therefore, synthetic observations had to be used. The synthetic truth model had a very high spatial resolution for the simulation of the Mølleåen River with river cross sections spaced 10 m from each other. The synthetic truth observations used for calibration, i.e. discharge and water level, and the target predictions, i.e. GW-SW exchange flows, were obtained from this synthetic truth model. Synthetic discharge observations were obtained for the gauging station highlighted with a magenta triangle in Figure 4. Water level observations were extracted for the branch of the Mølleåen River, downstream of Lyngby Sø, as highlighted by the black rectangle in Figure 4. Discharge and water level observations were perturbed with noise to represent the uncertainty of the measuring system. Discharge observations were corrupted with a 10%

Gaussian Noise to represent the inaccuracy of the rating-curve (Domeneghetti *et al.*, 2012); while water level observations were perturbed with 5 cm Gaussian noise to include the measuring errors of the UAV-borne ranging system (Bandini *et al.*, 2017).

The synthetic truth observations of GW-SW exchange flow are shown in Figure 5. In MIKE SHE, exchange flow is the portion of river discharge that originates from both shallow and deep subsurface flow. The exchange flow is computed as the amount of water per unit of time normalized by the unit of length of river segment.

Figure 5 shows a spatial series of GW-SW exchange flow observations, averaged over the evaluation period of the simulation. Sharpness and reliability of the GW-SW estimates are evaluated for: i) the spatial series, averaged over the evaluation period, along the entire branch; ii) the time series in the specific river stretch highlighted with a blue rectangle in Figure 5. This river stretch was chosen because the direction of the exchange flow is not constant but changes over time.

2.2.3. Calibration methodology

Calibration was performed with DREAM (DiffeRential Evolution Adaptive Metropolis) algorithm (Vrugt *et al.*, 2008) using MATLAB software. DREAM has the ability to separate between the different sources of uncertainty: input, such as boundary and initial conditions, parameters and structural errors.

2.2.3.1 Calibration parameters

Since the integrated hydrological model has a very large number of parameters, a primary one-at-a-time (OAT) sensitivity test was necessary to reduce the effort required for calibration by focusing on the most sensitive parameters, i.e. parameters causing significant uncertainty in the water level of the branch shown in Figure 5. Local sensitivity analysis showed that the most sensitive parameters were i) parameters affecting the free overflow discharge through river structures ii) Manning number, and iii) datum of two river cross sections.

Figure 5 shows the two cross sections at which the uncertainty in geodetic datum causes most uncertainty in water level. A constant Manning number was assigned to the entire branch. In Mike 11 the open channel resistance is defined as the roughness coefficient “ K_s ” (Strickler-Manning coefficient), which is simply the inverse of the Manning coefficient (generally referred to as “ n ”). The coefficient K_s varies from 6 to $40 \frac{m^{\frac{1}{3}}}{s}$ for natural channels (Chow, 1959).

In addition, the river structures have an important function in regulating river water level and require detailed modelling to improve the simulation of the river dynamics. The free overflow factors showed a larger impact on the water slope than the coefficients determining the head loss across the river structures. In MIKE 11 the free (critical) overflow over a structure is affected by the free overflow factor, a_c , as shown in equation (1).

$$Q = a_c \cdot Q_c \quad (1)$$

In equation (1) the critical flow discharge, Q_c , is multiplied by the free overflow factor to obtain the actual discharge, Q . The factor a_c is generally set as one, but can assume different values to represent that actual river cross sections are generally irregular. Indeed, in case of non-uniform cross sections, water level is not horizontal and the velocity distribution is not

uniform. In case of non-parallel flow (curved streamlines) over the weir, e.g. sharp-crested weir, a value greater than one can be applied, while, in case of side effects in the proximity of the weir, a value less than one should be considered. The first six upstream structures in the branch of Figure 5 significantly affect water level, thus their overflow factors were chosen as calibration parameters.

Table displays the list the calibration parameters and their prior uncertainty ranges, initial values and boundary handling as discussed below.

The range of the parameters defines the feasible parameter space from which the initial state of each of the chains is drawn with Latin hypercube sampling. Indeed, the DREAM algorithm runs multiple different chains simultaneously for global exploration of the parameter space (Beven, 2008). The total number of model runs is equal to $N \cdot T$, where N is the number of chains and T is the number of generations. For this model calibration, we use a number of chains, N , equal to the number of parameters (9) and a number of generations, T , equal to 900. Thus, the total number of model runs was 8100.

Vrugt (2016) states that during chain evolution, the bound has to be actively enforced, since candidate points can fall outside the hypercube defined by the defined range, albeit the initial state of each chain is within bounded search domain. In this case, when a proposal falls outside the range, it is reflected backwards into the parameter space with an "amount" equal to the boundary violation.

2.2.3.2 Discharge calibration (DC)

The vector of residuals, e_i , can be computed by subtracting the synthetic truth discharge, \hat{Q}_i , from the model-simulated discharge, Q_i , for each of the N observations. For discharge, 82

synthetic daily consecutive observations were considered for the in-situ discharge measuring station represented by a magenta triangle in Figure 4.

For this case study, we assume that initial conditions do not affect the model outputs, since a long warm-up period (5 years) is used to decrease the sensitivity to state-value initialization. Moreover, boundary conditions do not influence the outputs, since the synthetic truth model and the model to be calibrated share the same forcing boundary conditions. Assuming that the residuals are mutually independent and Gaussian distributed, the posterior pdf of the parameters, $p(\theta/\hat{Q})$, is described by equation (2).

$$p(\theta/\hat{Q}) = p(\theta) \cdot \prod_{i=1}^N \frac{1}{\sqrt{\sigma_i^2 \cdot 2\pi}} \cdot \exp\left(-\frac{e_i^2}{2\sigma_i^2}\right) \quad (2)$$

In equation (2) $p(\theta)$ is the prior distribution of the model parameters, σ_i is an estimate of the standard deviation of the i^{th} measurement.

However, it was more convenient to maximize the natural logarithm of the likelihood function shown in (3).

$$\ell(\theta/\hat{Q}) = -\frac{N}{2} \cdot \ln(2\pi) - \sum \ln(\sigma_i) - \frac{1}{2} \cdot \sum \left(\frac{\varepsilon_i}{\sigma_i}\right)^2 \quad (3)$$

The logarithm is a monotonically increasing function; thus, it achieves its maximum value at the same points as the likelihood function itself.

2.2.3.3 *Discharge and water level calibration (DWLC)*

In this case, the model is calibrated against both discharge and spatially distributed water level observations. For discharge, 82 daily observations, identical to the ones in the first calibration methodology, were used. A total number of 69 water level observations were retrieved from the high-resolution model. These synthetic observations were extracted during three different time steps, in each of the 23 simulated river cross sections that define the river geometry of the branch flowing from Lyngby Sø to the sea.

When both discharge and water level observations are available, the assumption that the residuals are still mutually independent and Gaussian distributed can be formulated. Thus the vector of the residual, e_i , is obtained by subtracting each of the model responses, Y_i , from the observations, \hat{Y}_i , which consists of either discharge or water level. The variable N in this calibration case is equal to 151: the total amount of discharge and water level observations. Discharge and water level residuals, as shown in (3), are normalized by dividing by the variance. Different weights were tested for the two different hydrological datasets, however they did not significantly improve the posterior distribution of the model parameters. Thus, uniform weights were adopted.

2.2.3.4 *Predictive uncertainty of the model response*

The convergence of each Markov chain to a stationary distribution was evaluated using Gelman-Rubin R-statistic (Gelman and Rubin, 2007). Then, from the ensemble obtained after convergence, the 2.5% and 97.5% percentiles of each of the models outputs are computed by analysing the distribution of model responses. The generated model response distribution only includes the uncertainty in the parameters. To include also the other error sources, such

as uncertainty in forcing conditions, measurement errors and model structure, the residual error is added to the model prediction for each MCMC (Markov Chain Monte Carlo) sample. To implement that, first, the RMSE (Root Mean Square Error) of the maximum a posteriori estimation (MAP) is computed. For the calibration against discharge and water level, two different RMSE values were computed, one value for water level and one value for discharge, since the assumption of homoscedasticity cannot be formulated. A residual error that is normally distributed, with zero mean and standard deviation equal to RMSE of the MAP, is added to each posterior estimation of the model response. When DWLC is performed, residual errors are added to the two different model outputs (discharge and water level), taking into account the two different RMSE values. Then, the 95% posterior confidence interval due to the total uncertainty can be computed.

In addition to investigating the model response in terms of discharge and water level, the effectiveness of the two calibration methodologies is evaluated by investigating the model estimation of the GW-SW interaction. GW-SW exchange flow observations are compared with synthetic truth simulation outputs. Sharpness and reliability of estimates are assessed with multiple scoring functions and statistics.

3. Results & Discussion

UAV-borne water level observations were retrieved in a specific river stretch. Then a synthetic study is conducted to analyze the potential of these spatially distributed water level observations in hydrological modelling.

3.1. UAV-borne water level observations

Figure 3 showed a water level that decreases by ca. 10 cm in the surveyed stretch. Despite local inaccuracies in water level determination, the water profile can be obtained with high spatial resolution and an accuracy of few cm. This slope is in agreement with the in-situ ground truth measurements conducted in that stretch, which were retrieved with the RTK GNSS station. UAV-borne observations showed an RMSE of ca. 4.1 cm, an MAE (mean absolute error) of 3.8 cm and an MBE (mean bias error) of 3.4 cm in the surveyed stretch.

The water level observations reported in this paper are presented as an integration of ground-based and UAV-borne observations retrieved in multiple flights. With the current technology and legislation, airborne observations of the entire branch could not be obtained. However, in the last years, navigation systems, such as GNSS receivers, inertial measurement units and autopilots, have advanced (Watts et al., 2012) and furthermore legislation is in continuous evolution. Therefore, with the state-of-art technology, a single flight covering the entire branch, exploiting BLOS (beyond line-of-sight) VTOL (vertical take-off and landing) UAVs (Watts *et al.*, 2012), is expected to be feasible.

Consistent and accurate UAV-borne measurements of floodplain and braided rivers can be used to inform calibrate and validate hydrological models, for example, to improve estimation of the amount of water stored in a river and floodplain and of its exchange with its main channel. Timing of the sampling can allow retrieval of measurements of water level and surface extension during extreme events, during which spaceborne observations are generally not available. Thus, we can expect that also flood mapping would be significantly improved by this innovative distributed water level datatype, because of the increasing potential of UAVs in the field of disaster prevention and mitigation (Zhang and Wu, 2014; Luo *et al.*,

2015). In this framework, we evaluated the potential of UAV-borne water level observations to improve estimation of GW-SW interaction.

3.2. DC vs DWLC results

Convergence was reached after around 5000 model runs in DC case, while was reached after around 4000 model runs in the DWLC case. Figure 6 (a) shows the a posteriori distribution of the model parameters after DC. While Figure 6 (b) shows the a posteriori distribution of the model parameters after DWLC.

Figure 6 (a) shows that only parameters a_{c1} , a_{c6} and K_s have distributions with a clear high-probability peak near the MAP. Instead, a_{c3} and g_{d1} have a distribution with a slightly higher marginal probability in a region far away from the MAP. Thus the maximum a posteriori probability is reached for values of a_{c3} and g_{d1} parameters that are not in the highest probability area. Parameter a_{c2} shows a distribution consisting of two disconnected modes with in between a region of low probability. Bimodality of this distribution slightly slowed convergence time. Indeed, covariance based MCMC methods suffer from multimodal target distribution since transition between probability regions is infrequent.

Figure 6 (b) shows that, for most of the parameters, the a posteriori distribution of the model parameters is better defined after DWLC. The distribution is unimodal, with a higher probability value close to the MAP for all the parameters, except for g_{d2} , which is a parameter determining the geodetic datum in one cross section.

The model response, in terms of discharge, is plotted in Figure 7 (a) for DC and in 7 (b) for DWLC.

Figure 7 shows that the model is able to simulate both the rising and falling limb of the streamflow hydrograph. Parameter uncertainty is less significant than remaining uncertainty (in this case model structure uncertainty and measurement errors). The impact of the MIKE11/SHE model structure on streamflow uncertainty has already been addressed in other studies (Butts *et al.*, 2004).

The parameter uncertainty range reliably predicts observations during low-flow days. However, during high-flow days, only few observations are contained within the parameter uncertainty range and even the light-grey remaining uncertainty range does not include some of the observations. However, the large spread in observations for the highest flows is also due to the synthetic measurement error for discharge, which is 10% of its absolute value.

Figure 7 (b) shows a parameter uncertainty that is narrower than in Figure 7 (a): i.e. sharpness in discharge estimation was improved.

Figure 8 represents the uncertainty in water depth estimations. Water depth is plotted instead of water level, since plotting water depth allows having a smaller range of values on the y-axis. Since the geometry of the 23 cross sections in the river branch can be accurately extracted from the in-situ bathymetric measurements, water depth can be directly computed from water level.

Figure 8 displays the uncertainty in water depth over the river branch for three different simulation time steps. Thus, the sharpness of the uncertainty range can be compared only for the dark grey range (parameter uncertainty). In the first time step, parameter uncertainty is significantly lower than in the remaining two time steps. This is also a consequence of the first time step being a dry day, with considerably lower water depth.

The total uncertainty range in water level can be computed only in case of DWLC, thus total uncertainty is shown only in the Figure 8 (b). Figure 8 (b) shows a parameter uncertainty significantly reduced after DWLC, especially in the second and third time step. Parameter uncertainty, at least in the upstream part of the river during the second and third time steps, comprises a large part of the total uncertainty. In this second and third time steps DWLC significantly reduces parameter uncertainty. However, in the first time step, the impact of the DWLC is lower.

3.3. Estimates of GW-SW exchange flow

Figure 9 shows the time series of exchange flow in the selected river stretch. Figure 9 clearly shows that, after DWLC, the width of the confidence interval significantly decreases and the 50th percentile becomes a better estimator of the observations.

Figure 10 displays the spatial series of exchange flow in the entire river branch, averaged over the evaluation period of the simulation.

Figure 10 shows that DWLC has a positive effect in both decreasing uncertainty and improving reliability of the estimates. For instance, the uncertainty range of estimates is able to include the observations around chainage $1.66 \cdot 10^4$ m only after the second calibration.

Statistics are reported in Table to compare the two calibration methodologies.

Regarding the time series of the exchange flow (Figure 9), the improvement in the sharpness and reliability is clearly highlighted by the ISS. Indeed, the ISS rewards narrow confidence intervals and penalizes confidence intervals which do not include observations. The ISS is approximately halved between the two calibrations, i.e. the model has improved sharpness without losing reliability. This is also confirmed by the approximately halved sharpness index, without any significant loss in coverage. Sharpness index represents the average width of the confidence interval, thus a decrease in sharpness index indicates narrower confidence bounds (i.e. sharpest estimates). The CRPS shows a radical improvement between the two calibrations. The CRPS is a generalization of the mean absolute error for probabilistic estimates: lower CRPS values indicate that the forecasts, expressed as probability distributions, are matching observed outcomes. The binary outcome of the mutually exclusive prediction of the exchange flow direction, i.e. gaining or losing stream, was evaluated using the Brier Score. The Brier score showed an improvement in predicting the direction of the exchange flow after DWLC.

Regarding the spatial series of the exchange flow (Figure 10), sharpness index improves of ca. 50%, but coverage results are worse in the second calibration methodology. However, that should not be interpreted as a loss of reliability. Some observations are only slightly outside the uncertainty range after the second calibration. However, observations with large prediction errors after DC are correctly estimated after DWLC. The decrease in RMSE and CRPS confirms this assumption. The Brier score shows a significant improvement.

4. Conclusions

UAVs can retrieve water level of small rivers at a spatial resolution of ca. 1 metre and with an accuracy of 5-7 cm. Furthermore, narrow rivers (few meters wide) can be accurately monitored.

The potential of this new datatype in improving river hydrological models was explored by presenting a synthetic study and investigating the improvement in the estimates of GW-SW interaction. The synthetic study consisted of the following steps:

- i. An integrated hydrological model, MIKE 11-MIKE SHE model, was used to simulate a small river and its catchment.
- ii. Calibration against highly spatially distributed water levels, has been compared with calibration only against discharge observations.
- iii. GW-SW estimates of the hydrological model were compared after the two calibration methodologies. The sharpness and reliability of GW-SW estimates were investigated for i) the spatial series of GW-SW exchange flow in the entire investigated river branch ii) the time series of GW-SW exchange flow time in a selected river stretch.

After the calibration against UAV-borne synthetic water level observations, sharpness and reliability of the estimates substantially improved:

- Sharpness improves of ca. 50%
- The Brier Score shows a significant improvement. This suggests that the model is better at predicting the direction of the exchange flow.

- The RMSE substantially decreases. The RMSE is reduced to ca. 25% of the initial RMSE, which was computed after discharge calibration, for the time series of GW-SW interaction. Thus, the model predictions are generally more reliable after water level calibration. This was confirmed also by other statistics such as the CRPS.

Acknowledgments

The authors would like to thank the Innovation Fund Denmark for financial support via the project “Smart UAV” [125-2013-5]. The authors are grateful to J. A. Vrugt for providing the latest version of the DREAM algorithm.

- Alsdorf DE, Rodriguez E, Lettenmaier DP. 2007. Measuring surface water from space. *Reviews of Geophysics* **45** (2): 1–24 DOI: 10.1029/2006RG000197.1
- Baghdadi N, Lemarquand N, Abdallah H, Bailly JS. 2011. The Relevance of GLAS/ICESat Elevation Data for the Monitoring of River Networks. *Remote Sensing* **3** (12): 708–720 DOI: 10.3390/rs3040708
- Bandini F, Jakobsen J, Olesen D, Reyna-Gutierrez JA, Bauer-Gottwein P. 2017. Measuring water level in rivers and lakes from lightweight Unmanned Aerial Vehicles. *Journal of Hydrology* **548**: 237–250 DOI: 10.1016/j.jhydrol.2017.02.038
- Beven KJ. 2008. Comment on ‘Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?’ by Jasper A. Vrugt, Cajo J. F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson. *Stochastic Environmental Research and Risk Assessment* **23** (7): 1059–1060 DOI: 10.1007/s00477-008-0283-x
- Breinholt A, Møller JK, Madsen H, Mikkelsen PS. 2012. A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation - distinguishing simulation and prediction. *Journal of Hydrology* **472-473**: 36–52 DOI: 10.1016/j.jhydrol.2012.09.014
- Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78** (1): 1–3 DOI: 10.1126/science.27.693.594
- Butts MB, Payne JT, Kristensen M, Madsen H. 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* **298** (1-4): 242–266 DOI: 10.1016/j.jhydrol.2004.03.042
- Chow VT. 1959. Open Channel Hydraulics. *McGraw-Hill Book Company*: 728 DOI: ISBN 07-010776-9
- Domeneghetti A, Castellarin A, Brath A. 2012. Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrology and Earth System Sciences* **16** (4): 1191–1202 DOI: 10.5194/hess-16-1191-2012
- Domeneghetti A, Castellarin A, Tarpanelli A, Moramarco T. 2015. Investigating the uncertainty of satellite altimetry products for hydrodynamic modelling. *Hydrological Processes* **29** (23): 4908–4918 DOI: 10.1002/hyp.10507
- Gelman A, Rubin DB. 2007. Inference from Iterative Simulation Using Multiple Sequences. *7* (4): 457–472
- Giordan D, Franzi L, Rinaldi S, Dutto F, Baldo M, Allasia P. 2016. Multi-temporal image analysis for river reach morphological changes identification. *Geophysical Research Abstracts EGU General Assembly* **18** (23): 2016–13142
- Gneiting T, Raftery AE. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102** (477): 359–378 DOI: 10.1198/0162145060000001437
- Gneiting T, Raftery AE, Westveld AH, Goldman T. 2005. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review* **133** (5): 1098–1118 DOI: 10.1175/MWR2904.1
- Graham DN, Butts MB. 2005. Flexible Integrated Watershed Modeling with MIKE SHE. In *Watershed Models* 245–272. DOI: 10.1201/9781420037432.ch10

- Hall AC, Schumann GJP, Bamber JL, Bates PD, Trigg MA. 2012. Geodetic corrections to Amazon River water level gauges using ICESat altimetry. *Water Resources Research* **48** (6) DOI: 10.1029/2011WR010895
- Hopkinson C, Crasto N, Marsh P, Forbes D, Lesack L. 2011. Investigating the spatial distribution of water levels in the Mackenzie Delta using airborne LiDAR. *Hydrological Processes* **25** (19): 2995–3011 DOI: 10.1002/hyp.8167
- Hughes JD, Liu J. 2008. MIKE SHE: Software for Integrated Surface Water/Ground Water Modeling. *Ground Water* **46** (6): 797–802 DOI: 10.1111/j.1745-6584.2008.00500.x
- Kleinherenbrink M, Ditmar PG, Lindenbergh RC. 2014. Retracking Cryosat data in the SARIn mode and robust lake level extraction. *Remote Sensing of Environment* **152**: 38–50 DOI: 10.1016/j.rse.2014.05.014
- Lawford R, Strauch A, Toll D, Fekete B, Cripe D. 2013. Earth observations for global water security. *Current Opinion in Environmental Sustainability* **5** (6): 633–643 DOI: 10.1016/j.cosust.2013.11.009
- Legresy B, Papa F, Remy F, Vinay G, Van Den Bosch M, Zanife OZ. 2005. ENVISAT radar altimeter measurements over continental surfaces and ice caps using the ICE-2 retracking algorithm. *Remote Sensing of Environment* **95** (2): 150–163 DOI: 10.1016/j.rse.2004.11.018
- Luo C, Nightingale J, Asemota E, Grecos C. 2015. A UAV-Cloud System for Disaster Sensing Applications. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)* IEEE; 1–5. DOI: 10.1109/VTCSpring.2015.7145656
- Maillard P, Bercher N, Calmant S. 2015. New processing approaches on the retrieval of water levels in Envisat and SARAL radar altimetry over rivers: A case study of the São Francisco River, Brazil. *Remote Sensing of Environment* **156**: 226–241 DOI: 10.1016/j.rse.2014.09.027
- Mason DC, Horritt MS, Hunter NM, Bates PD. 2007. Use of fused airborne scanning laser altimetry and digital map data for urban flood modelling. *Hydrological Processes* **21** (11): 1436–1447 DOI: 10.1002/hyp.6343
- Munz M, Krause S, Tecklenburg C, Binley A. 2011. Reducing monitoring gaps at the aquifer-river interface by modelling groundwater-surface water exchange flow patterns. *Hydrological Processes* **25** (23): 3547–3562 DOI: 10.1002/hyp.8080
- Niedzielski T, Witek M, Spallek W. 2016. Observing river stages using unmanned aerial vehicles. *Hydrol. Earth Syst. Sci* **20**: 3193–3205 DOI: 10.5194/hess-20-3193-2016
- O'Loughlin FE, Neal J, Yamazaki D, Bates PD. 2016. ICESat-derived inland water surface spot heights. *Water Resources Research* **52** (4): 3276–3284 DOI: 10.1002/2015WR018237
- Phan VH, Lindenbergh R, Menenti M. 2012. ICESat derived elevation changes of Tibetan lakes between 2003 and 2009. *International Journal of Applied Earth Observation and Geoinformation* **17** (1): 12–22 DOI: 10.1016/j.jag.2011.09.015
- Ritter A, Muñoz-Carpena R. 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology* **480**: 33–45 DOI: 10.1016/j.jhydro.2012.12.004
- Schumann G, Matgen P, Cutler MEJEJ, Black A, Hoffmann L, Pfister L. 2008. Comparison

- of remotely sensed water stages from LiDAR, topographic contours and SRTM. *ISPRS Journal of Photogrammetry and Remote Sensing* **63** (3): 283–296 DOI: 10.1016/j.isprsjprs.2007.09.004
- Schumann GJ-P, Domeneghetti A. 2016. Exploiting the proliferation of current and future satellite observations of rivers. *Hydrological Processes* **30** (16): 2891–2896 DOI: 10.1002/hyp.10825
- Sole A, Zuccaro G. 2003. Comparison Between Open Channel Flow Models In Natural Rivers. *WIT Transactions on Ecology and the Environment* **60** DOI: 10.2495/RM030071
- Villadsen H, Andersen OB, Stenseng L, Nielsen K, Knudsen P. 2015. CryoSat-2 altimetry for river level monitoring — Evaluation in the Ganges–Brahmaputra River basin. *Remote Sensing of Environment* **168**: 80–89 DOI: 10.1016/j.rse.2015.05.025
- Vrugt J a., ter Braak CJF, Clark MP, Hyman JM, Robinson B a. 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* **44** (12): W00B09 DOI: 10.1029/2007WR006720
- Vrugt JA. 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software* **75**: 273–316 DOI: 10.1016/j.envsoft.2015.08.013
- Watts AC, Ambrosia VG, Hinkley EA. 2012. Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use. *Remote Sensing* **4(6)**: 1671–1692 DOI: 10.3390/rs4061671
- Wingham DJ, Francis CR, Baker S, Bouzinac C, Brockley D, Cullen R, de Chateau-Thierry P, Laxon SW, Mallow U, Mavrocordatos C, et al. 2006. CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields. *Advances in Space Research* **37** (4): 841–871 DOI: 10.1016/j.asr.2005.07.027
- Yamazaki D, Lee H, Alsdorf DE, Dutra E, Kim H, Kanae S, Oki T. 2012. Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resources Research* **48** (9) DOI: 10.1029/2012WR011869
- Yan K, Di Baldassarre G, Solomatine DP, Schumann GJP. 2015. A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. *Hydrological Processes* **29** (15): 3368–3387 DOI: 10.1002/hyp.10449
- Zhang W, Wu J. 2014. To Explore the UAV Application in Disaster Prevention and Reduction. *Applied Mechanics and Materials* **590**: 609–612 DOI: 10.4028/www.scientific.net/AMM.590.609

Table 1. Definition of the parameters and of their location, range and initial sampling distribution

Parameter symbol	Parameter description	Location in the branch shown in Figure 5	Range	Initial sampling distribution	Boundary handling
$a_{c1}, a_{c2}, a_{c3}, a_{c4}, a_{c5}, a_{c6},$	Free overflow factor	Six upstream structures	$0.3 < a_c < 1.8$	Latin hypercube	reflection
g_{d1}	Geodetic datum	Upstream cross section	$-0.3 < g_{d1} < 0.3$	Latin hypercube	reflection
g_{d2}	Geodetic datum	Downstream cross section	$-0.15 < g_{d2} < 0.15$	Latin hypercube	reflection
K_s	Strickler-Manning coefficient	Entire branch	$6 < K_s < 40$	Latin hypercube	reflection

Table II. Statistics computed for GW-SW exchange flows to compare the two different calibration methodologies.

Statistic	Time series of exchange flow in the selected river stretch		Spatial series of exchange flow in the entire branch		Reference for specific statistic
	DC	DWLC	DC	DWLC	
ISS [m ³ /(s·m)]	$1.81 \cdot 10^{-5}$	$7.77 \cdot 10^{-6}$	$7.89 \cdot 10^{-5}$	$1.17 \cdot 10^{-4}$	(Gneiting and Raftery, 2007)
CRPS [m ³ /(s·m)]	$2.65 \cdot 10^{-8}$	$9.36 \cdot 10^{-9}$	$1.57 \cdot 10^{-7}$	$1.21 \cdot 10^{-7}$	(Gneiting <i>et al.</i> , 2005)
RMSE [m ³ /(s·m)]	$8.39 \cdot 10^{-8}$	$1.95 \cdot 10^{-8}$	$2.97 \cdot 10^{-7}$	$2.77 \cdot 10^{-7}$	(Ritter and Muñoz- Carpena, 2013)
Sharpness [m ³ /(s·m)]	$2.88 \cdot 10^{-7}$	$1.21 \cdot 10^{-7}$	$5.71 \cdot 10^{-7}$	$1.97 \cdot 10^{-7}$	(Breinholt <i>et al.</i> , 2012)
Coverage [%]	100	98.41	47.82	17.39	
BS [-]	2.36	0.66	$2.83 \cdot 10^{-4}$	$1.88 \cdot 10^{-6}$	(Brier, 1950)

ISS is Interval Skill Score, CRPS is Continuous Ranked Probability Score, RMSE is Root Mean Square Error, BS is Brier score. Statistics were evaluated at the 0.05 significance level.

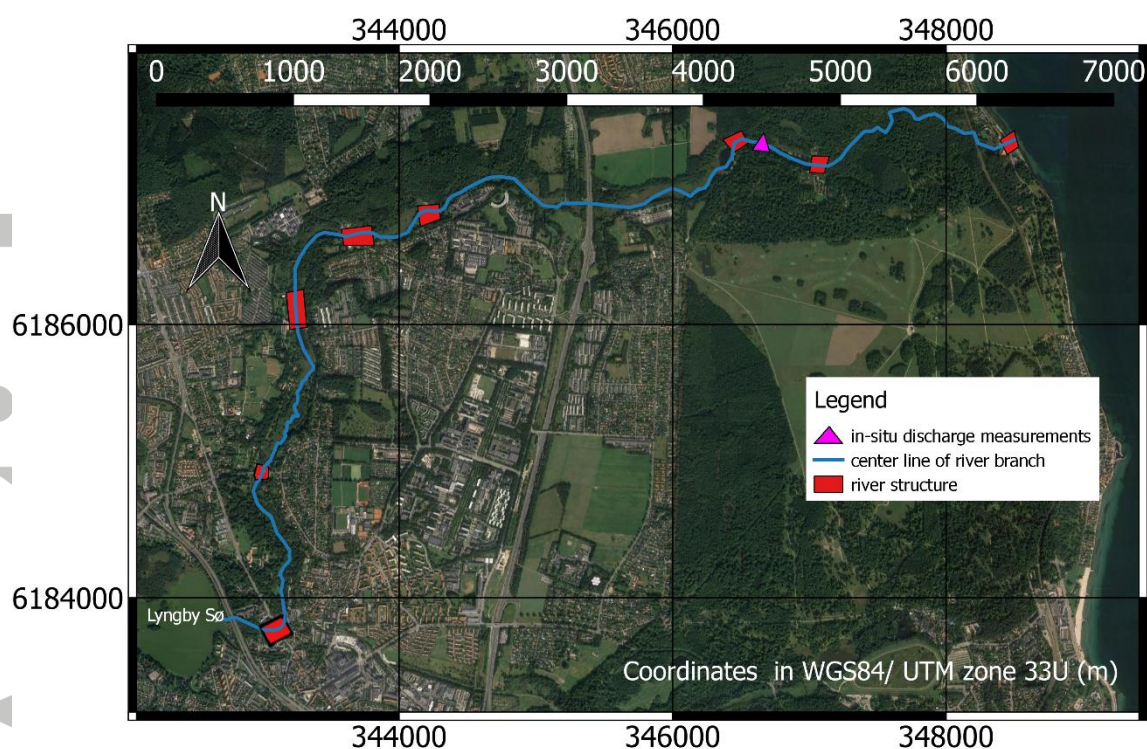


Figure 1. Branch of the Mølleåen River in which water level observations were obtained. The river branch includes eight structures (weirs). Along this branch, there are in-situ stations measuring discharge, but none of these are measuring water levels in real-time.

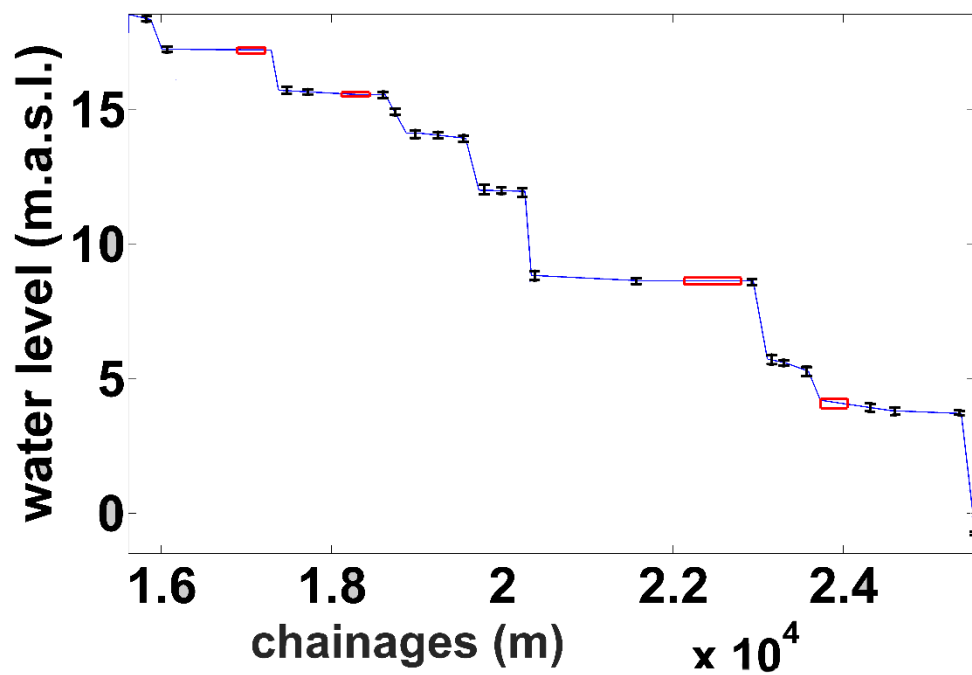


Figure 2. Water level (meters above mean sea level) is plotted against river chainage. UAV-borne observations are shown with a red rectangle, the height of which shows the average standard deviation. Black error bars shows standard deviation of ground-based observations.

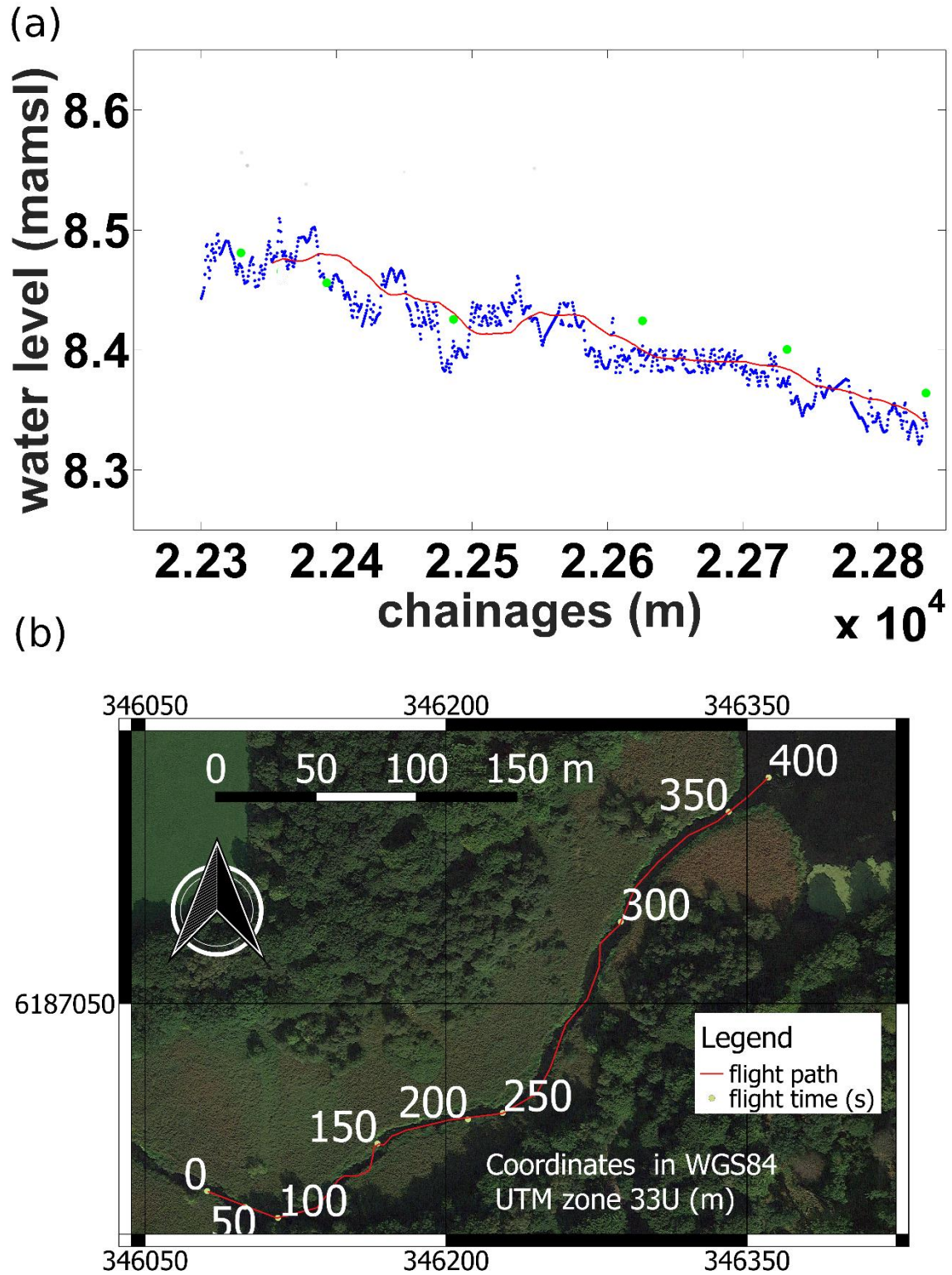


Figure 3. UAV measurements for a specific river reach. (a) Blue dots represent UAV-borne water level observations; red line shows the filtered UAV-borne water level observations; green dots represent in-situ observations retrieved with GNSS rover station. (b) Flight path above the river reach and flight way-points labelled with flight time elapsed since take-off.

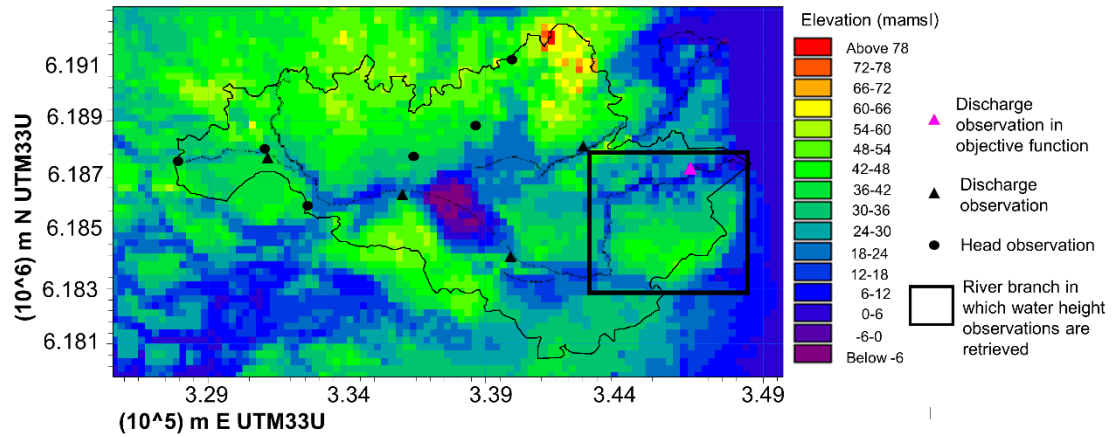


Figure 4. The entire catchment area, including the Mølleåen River, the lakes and other minor creeks. The integrated hydrological model was initially calibrated against the aquifer head observations (black dots) and the discharge observations (black triangles). A calibration of the river model, using DREAM algorithm, was performed against water levels and discharge. Water levels were retrieved in the branch highlighted with the black rectangle. While synthetic discharge observations were extracted from the gaging station highlighted with the magenta triangle.

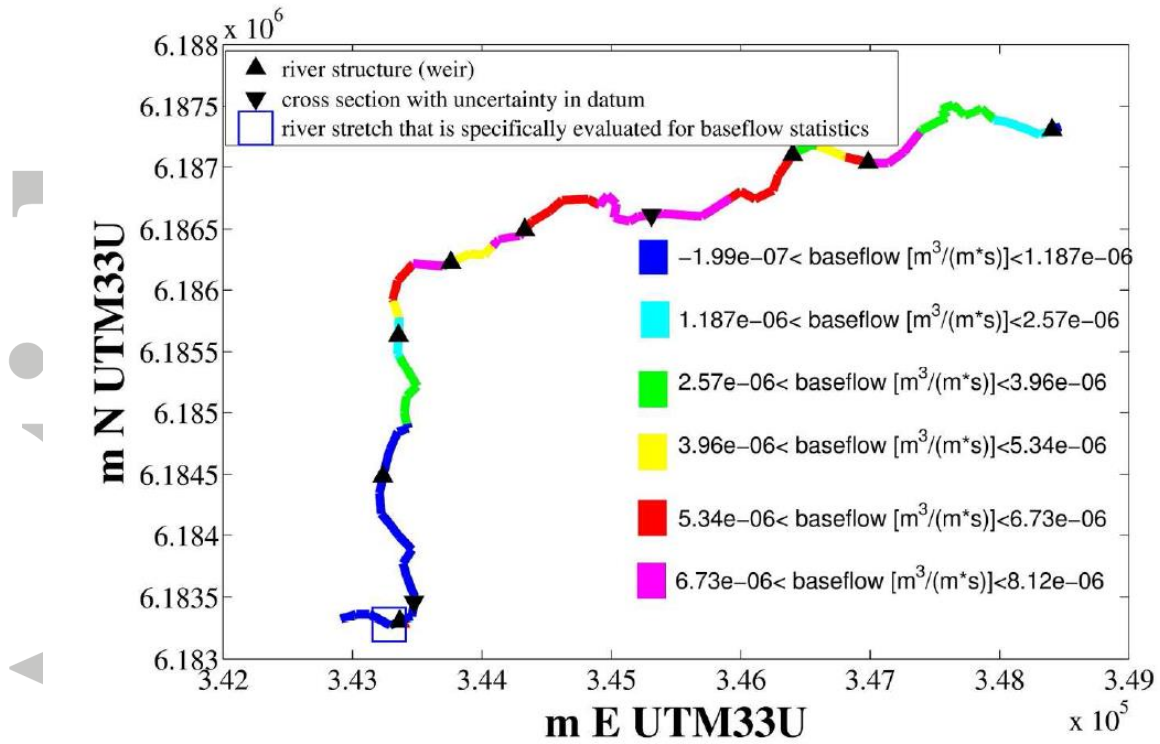


Figure 5. Synthetic GW-SW exchange flow observations, averaged over the evaluation period. The eight river structures are indicated with black up-pointing triangles. Black down-pointing triangles highlight the two cross sections of which the geodetic datum will be used as calibration parameter.

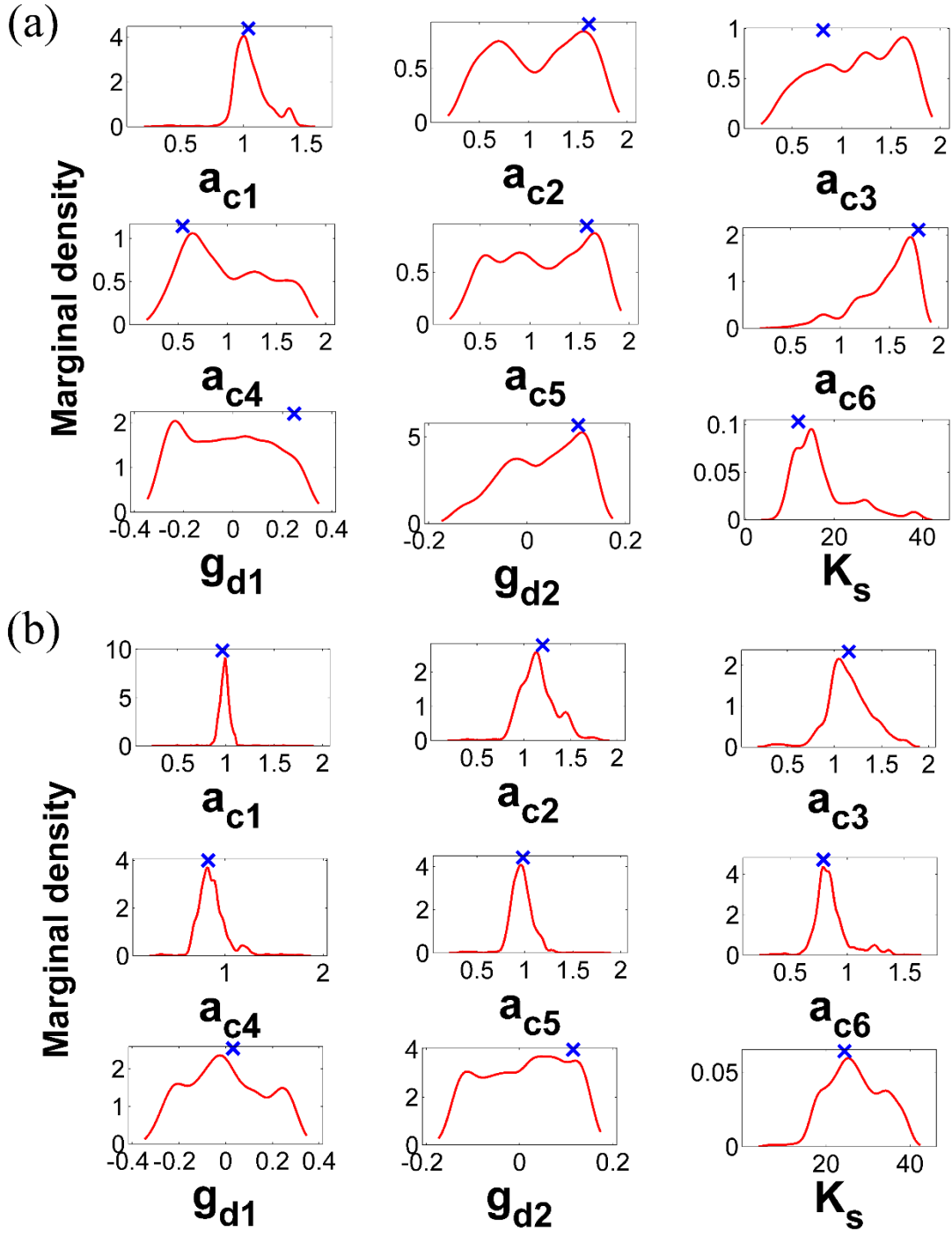


Figure 6. Marginal posterior density of model parameters after DC (a) and after DWLC (b). Blues cross indicates the parameter value identified by the MAP.

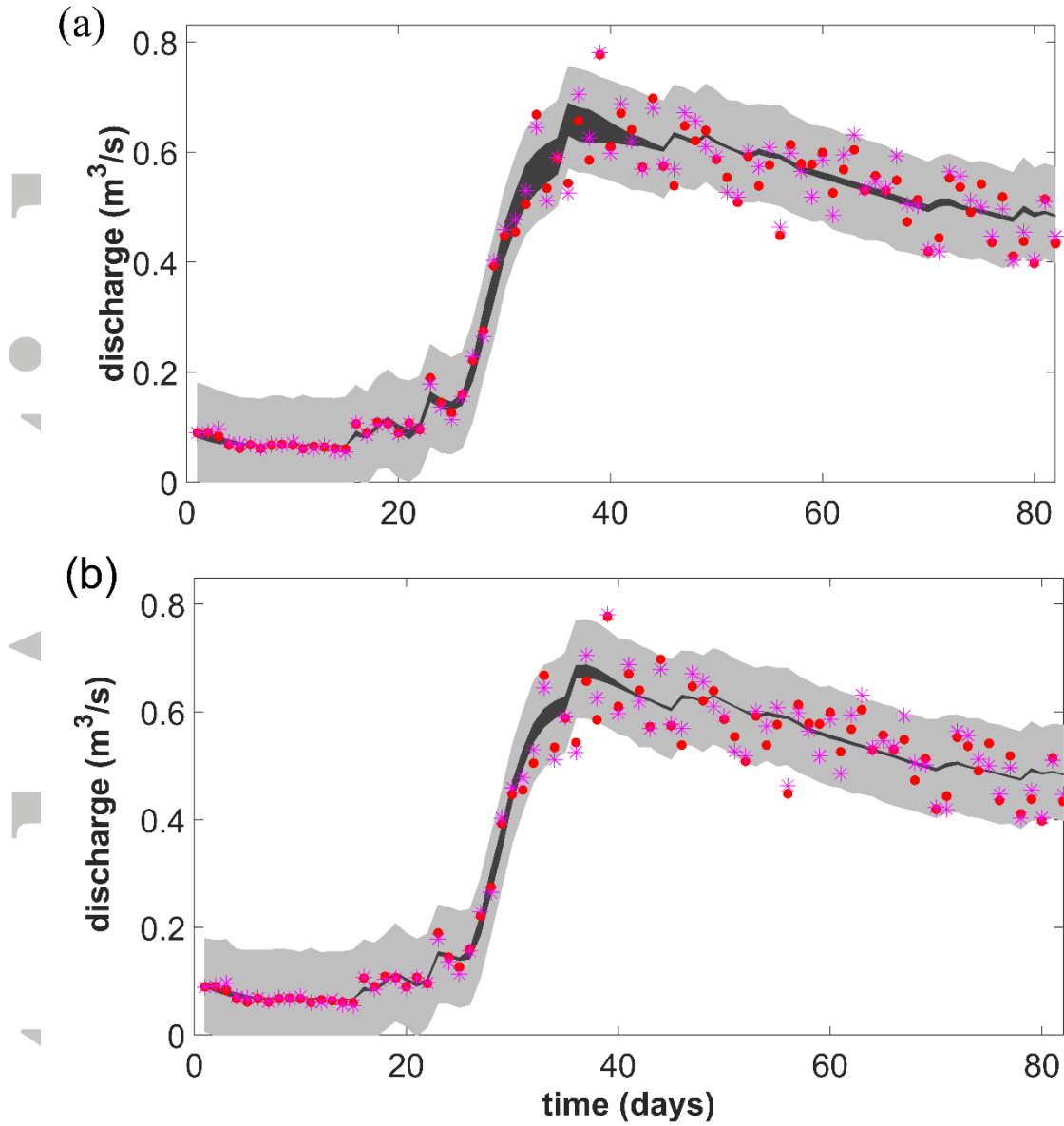


Figure 7: Posterior simulation of the uncertainty range after DC (a) and after DWLC (b). Red dots are the perturbed synthetic discharge observations used in the calibration. Magenta asterisks are the unperturbed synthetic observations. Day numeration: day 1 is 09-09-1999. The dark grey color is the 95% confidence intervals of the output prediction due to parameter uncertainty. Light grey region represents the remaining 95% prediction uncertainty.

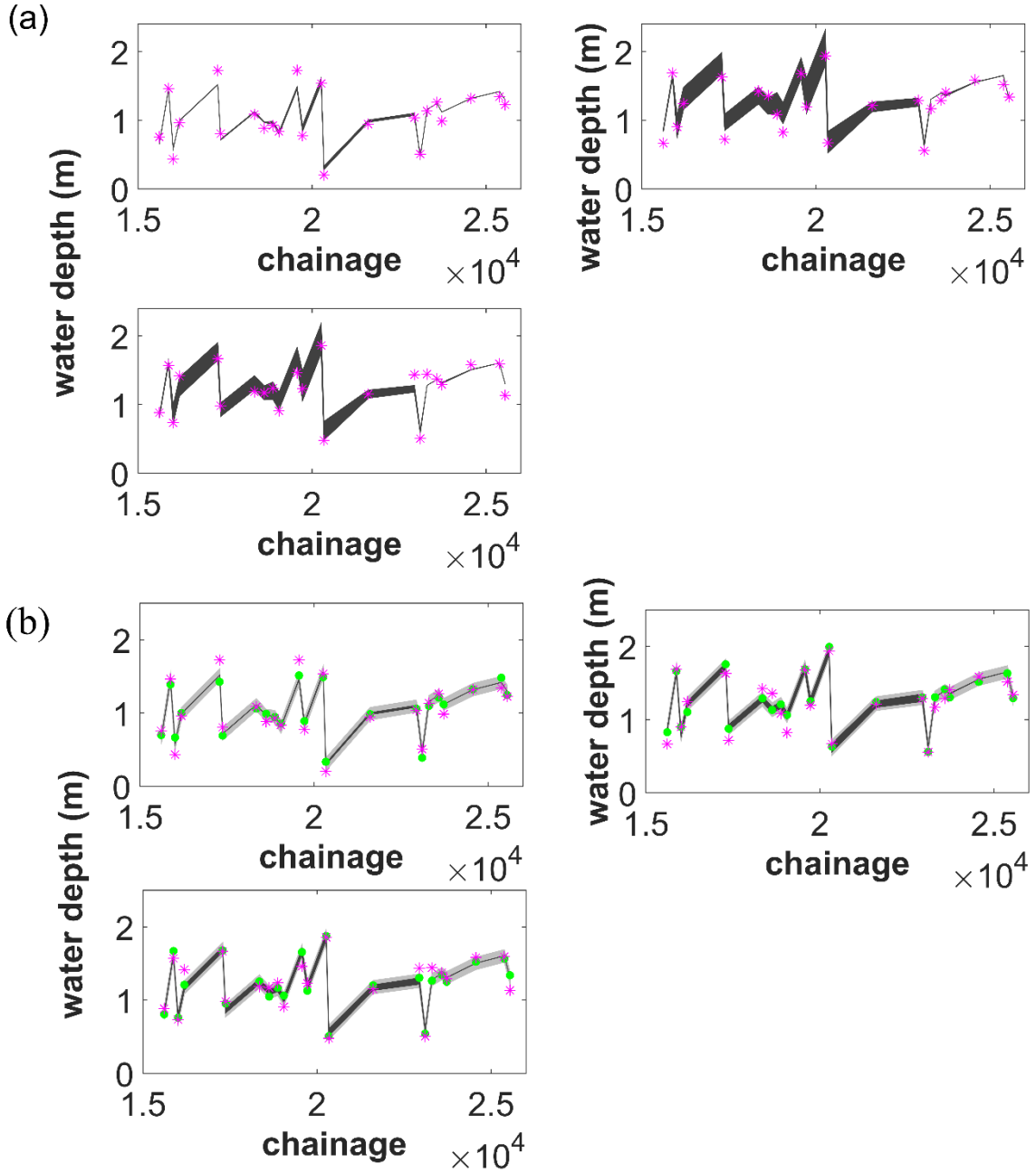


Figure 8: 95% Posterior simulation uncertainty ranges for water depth for the three different time steps, after DC (a) and after DWLC (b). For both (a) and (b), the top left panel represents day 7 (15-09-1999), the top right panel day 49 (27-10-1999) and the bottom panel day 82 (29-11-1999). The green dots are the perturbed synthetic water depth observations used for the calibration. Magenta asterisks are the unperturbed synthetic observations. The dark grey color is the 95% confidence intervals of the output prediction due to parameter uncertainty. Light grey region represents the remaining 95% prediction uncertainty.

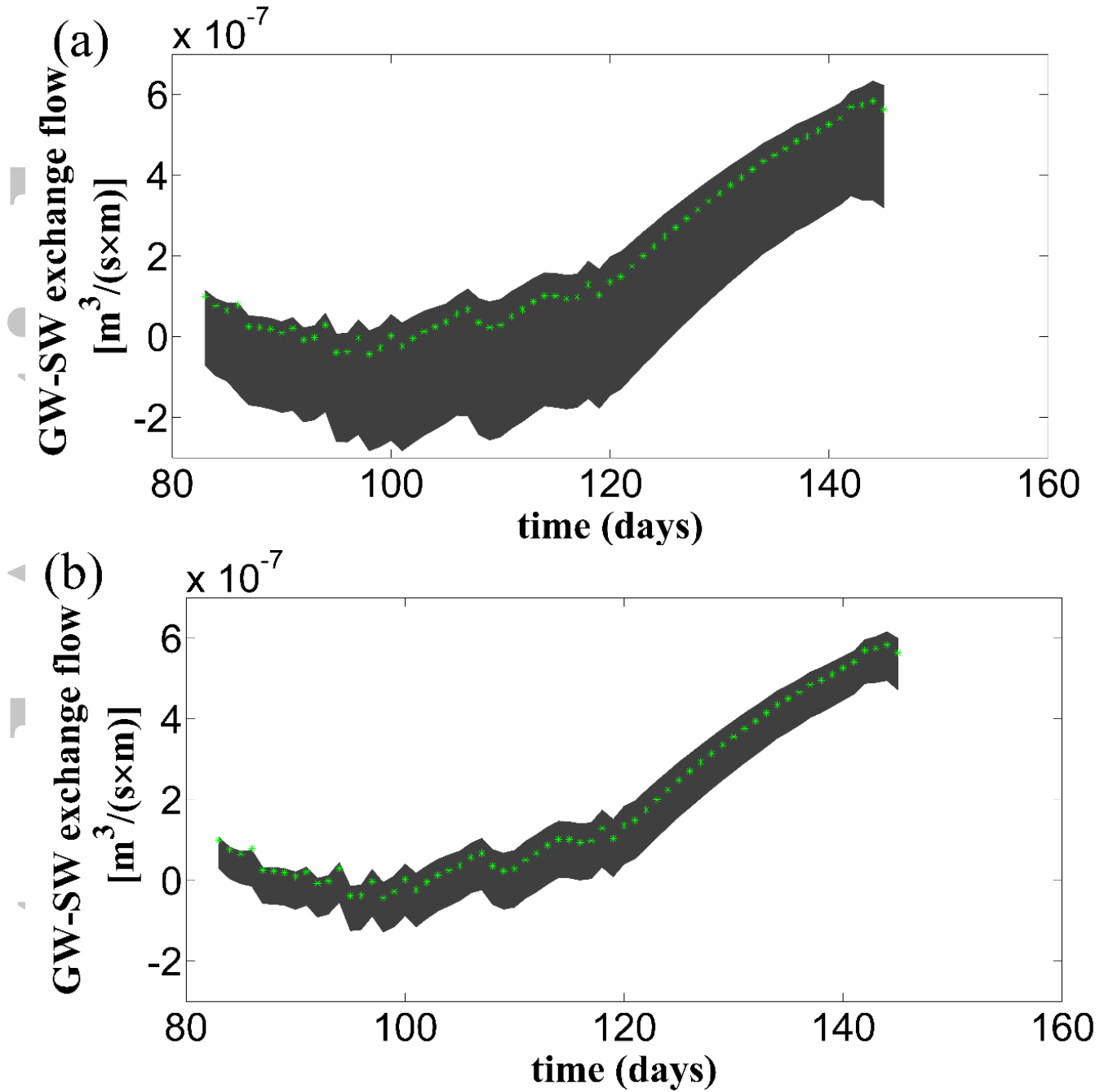


Figure 9. 95% Posterior simulation uncertainty ranges for exchange flow (m²/s) for specific river reach. (a) is after DC and (b) is after DWLC. Day numeration: day 83 is 30-11-1999. The green dots are the synthetic observations of exchange flow. The dark grey color is the 95% confidence intervals due to parameter uncertainty.

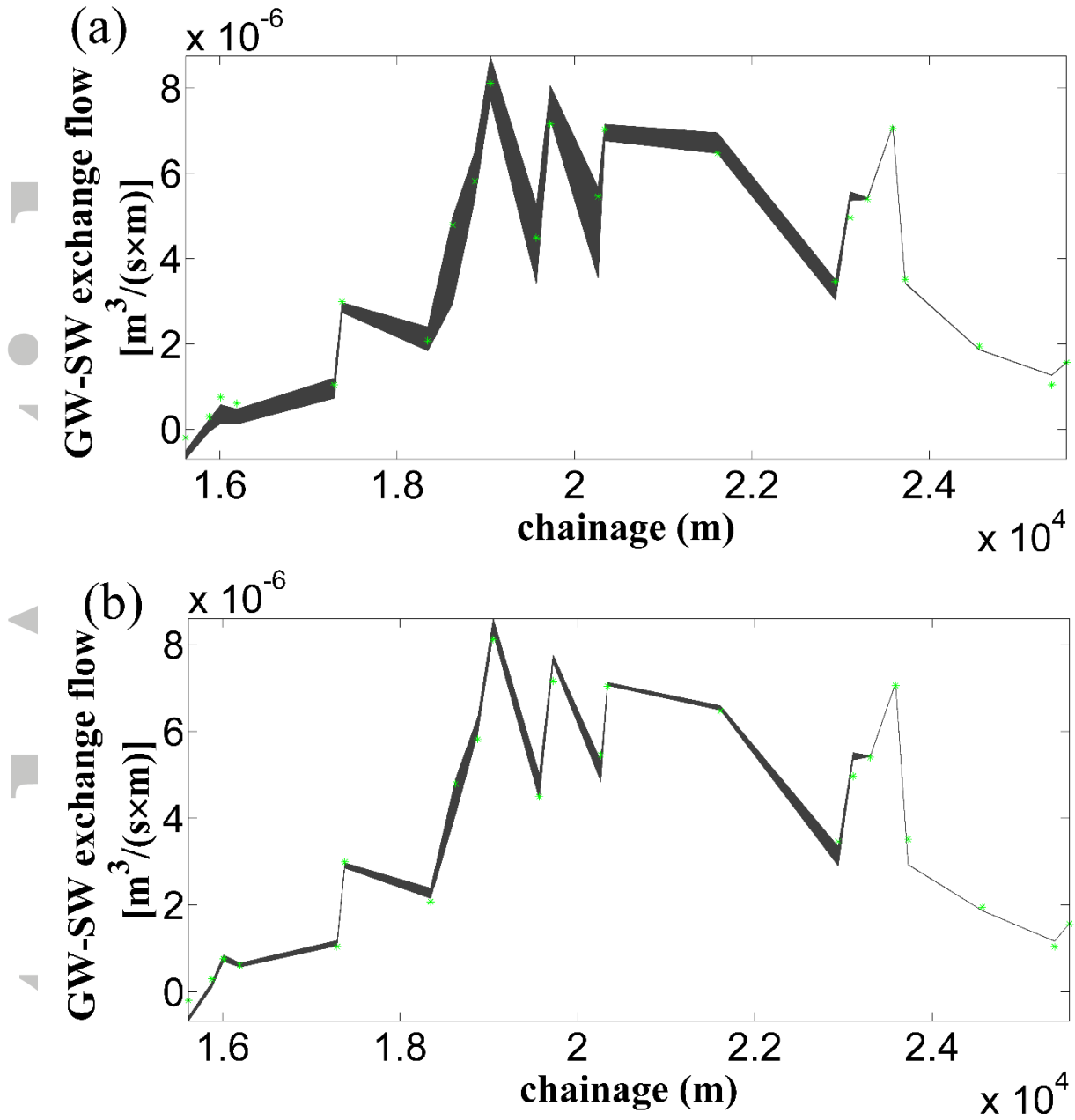


Figure 10. 95% Posterior simulation uncertainty ranges for exchange flow (m^2/s) for entire branch averaged over the evaluation period (from 30-11-1999 to 1-2-2000). (a) is after DC and (b) is after DWLC. The green dots are the synthetic observations of exchange flow for each of the cross sections. The dark grey color is the 95% confidence intervals due to parameter uncertainty.